Deux etudes en programmation non lineaire

J. Frederic Bonnans

IRIA

CENTRE DE ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France

Tél. : (1) 39 63 55 11

Rapports de Recherche

N° 548

DEUX ÉTUDES EN PROGRAMMATION NON LINÉAIRE

Joseph Frédéric BONNANS

Juillet 1986

DEUX ETUDES EN PROGRAMMATION NON LINEAIRE

Joseph Frédéric BONNANS#

RESUME

La première étude concerne le comportement des solutions locales lorsqu'on perturbe un problème d'optimisation non convexe. On y donne des conditions sous lesquelles une solution isolée bifurque en un nombre fini de solutions, qu'on peut effectivement calculer. La seconde étude établit l'augmentabilité et la pénalizabilité exacte au voisinage de solutions locales vérifiant des conditions suffisantes (faibles) du deuxième ordre.

ABSTRACT

The first study is concerned with the behaviour of local solutions of a perturbed non-convex optimization problem. Conditions are given that ensure that an isolated solution bifurcates into a finite number of solutions that can be effectively computed. The second study establishes the augmentability and the exact penalizability in the neighbourhood of local solutions satisfying some weak second-order sufficient conditions.

INRIA, Domaine de Voluceau, BP 105, Rocquencourt, 78153 Le Chesnay Cedex.

A SEMI-STRONG SUFFICIENCY CONDITION FOR OPTIMALITY IN NONCONVEX PROGRAMMING AND ITS CONNECTION TO THE PERTURBATION PROBLEM

Joseph Frédéric BONNANS#

ABSTRACT

What happens when a nonconvex program, having a local solution x^0 at which the gradients of the binding constraints are linearly independent, but without strict complementarity hypothesis, is perturbed ? Under a relatively weak second-order assumption (some non-negative second order terms are supposed to be strictly positive) the perturbed problem has, in the neighborhood of x^0, a finite number of local minima, situated on curves that are connected to some "pseudo-solutions" of the tangent quadratic problem.

INRIA, Domaine de Voluceau, BP 105, Rocquencourt, 78153 Le Chesnay Cedex, France.

I - INTRODUCTION

Let us consider for every $\varepsilon \in \mathbf{R}$ the following nonlinear program

$$\begin{aligned} & \text{Min } f^{\varepsilon}(x),\ x \in \mathbf{R}^n, \\ & \text{s.t. } g_i^{\varepsilon}(x) = 0,\ i = 1, \ldots, m, \\ & \qquad g_i^{\varepsilon}(x) \leq 0,\ i = m+1, \ldots, p, \end{aligned} \tag*{$(1.1)_\varepsilon$}$$

where f^{ε} and g_i^{ε}, $i = 1, \ldots, p$, are C^2 functions from $\mathbf{R}^n$ onto $\mathbf{R}$ continuous with respect to ε, such that their gradients ∇f^{ε} and ∇g_i^{ε} are C^1 with respect to (x,ε). We denote

$$\begin{aligned} & I^{\varepsilon}(x) = \{i = 1, \ldots, p\ ;\ g_i^{\varepsilon}(x) = 0\}, \\ & f(x) = f^0(x), \\ & g_i(x) = g_i^0(x),\ i = 1, \ldots, p. \end{aligned}$$

Let x^0 be a local solution of $(1.1)_0$ and suppose that

$$\nabla g_i(x^0),\ i = 1, \ldots, p, \text{ are linearly independent.} \tag{1.2}$$

Then there exists a unique Lagrange multiplier λ^0 in $\mathbf{R}^p$ such that

$$\begin{aligned} & \text{(i)} \quad \nabla f(x^0) + \sum_{i=1}^{p} \lambda_i^0 \nabla g_i(x^0) = 0, \\ & \text{(ii)} \quad g_i(x^0) = 0,\ i = 1, \ldots, m, \\ & \text{(iii)} \quad g_i(x^0) \leq 0,\ \lambda_i^0 \geq 0,\ \lambda_i^0 g_i(x^0) = 0,\ i = m+1, \ldots, p. \end{aligned} \tag{1.3}$$

As we are interested in what happens in the neighborhood of x^0 we may suppose, to simplify the study, that

$$g_i(x^0) = 0,\ i = 1, \ldots, p.$$

Let us denote

$$\begin{aligned} & I_0^0 = \{i = m+1, \ldots, p\ ;\ \lambda_i^0 = 0\}, \\ & H = \nabla^2 f(x^0) + \sum_{i=1}^{p} \lambda_i^0\, \nabla^2 g_i(x^0). \end{aligned}$$

A necessary condition for x^0 to be a local solution of $(1.1)_0$, called the <u>second order necessary condition</u>, is

$$\left.\begin{array}{l}\nabla g_i(x^0)^t d = 0,\ i \in (1,p) - I_0^0,\\[2ex] \nabla g_i(x^0)^t d \le 0,\ i \in I_0^0\end{array}\right\} \Rightarrow d^t\,Hd \ge 0. \qquad (1.4)$$

By changing an inequality into a strict inequality we obtain the <u>standard second-order sufficiency condition</u> :

$$\left.\begin{array}{l}d \neq 0,\\ \nabla g_i(x^0)^t d = 0,\ i \in (1,p) - I_0^0,\\ \nabla g_i(x^0)^t d \le 0,\ i \in I_0^0\end{array}\right\} \Rightarrow d^t\,Hd > 0. \qquad (1.5)$$

Condition (1.5) ensures that x^0 is a locally isolated solution of $(1.1)_0$. Together with (1.2), it implies the following stability result (see S.M. Robinson [3, Thm. 3.1 and Cor. 4.3]) :

<u>Theorem 1.1</u>

Let x^0 be a local solution of $(1.1)_0$ and λ^0 its Lagrange multiplier such that (1.2), (1.3) and (1.5) hold. Then there exists $a_1 > 0$ such that, for $|\varepsilon| < \varepsilon^0$, problem $(1.1)_\varepsilon$ possesses (at least) one local solution x^ε, with an associated Lagrange multiplier λ^ε, such that

$$||x^\varepsilon - x^0|| + ||\lambda^\varepsilon - \lambda^0|| \le a_1 |\varepsilon| . \qquad (1.6)$$

In addition, there exist $\varepsilon^0 > 0$, $\alpha > 0$, such that for $|\varepsilon| < \varepsilon^0$, any local solution x^ε such that $||x^\varepsilon - x^0|| < \alpha$ has an associated Lagrange multiplier λ^ε such that (1.6) holds. □

Stronger results have been obtained with the hypothesis that the following <u>strong second-order sufficiency condition</u> holds :

$$\left.\begin{array}{l}d \neq 0,\\[2ex] \nabla g_i(x^0)^t d = 0,\ i \in (1,p) - I_0^0\end{array}\right\} \Rightarrow d^t\,Hd > 0. \qquad (1.7)$$

We notice that (1.7) implies (1.5). Using (1.7), K. Jittorntrum [2] proved the

Theorem 1.2

Assume that the hypotheses of Theorem 1.1 hold, with (1.7) instead of (1.5). Then for ε small enough, there is a unique solution x^{ε} of $(1.1)_{\varepsilon}$ such that (1.6) holds, and the mapping $\varepsilon \to (x^{\varepsilon}, \lambda^{\varepsilon})$ has directional derivatives. □

If I_0^0 is empty, i.e. the strict complementarity conditions are fulfilled, (1.5) and (1.7) are equivalent and the mapping $(x^{\varepsilon}, \lambda^{\varepsilon})$ is C^1 (see A.V. Fiacco [1]). It is easy to see that if card $(I_0^0) = 1$, conditions (1.5) and (1.7) are also equivalent, but this is no longer true, in general, if card $(I_0^0) \geq 2$. The main drawback of the strong second-order sufficiency condition is that it assumes the strict positivity of an amount which can be strictly negative for an isolated minimum. Hence it is very strong. In contrast, we will use an hypothesis, in which we only assume the strict positivity of nonnegative amounts.

Definition 1.1

We will say that (I_1, I_2) is a pseudo-optimal subpartition of (1,p) (or shortly a pseudo-optimal subpartition) if

$$I_1 \supset (1,p) - I_0^0,$$

$$I_2 \subset (1,p) - I_1,$$

and

$$\left.\begin{array}{l} \nabla g_i(x^0)^t d = 0, \ i \in I_1, \\ \nabla g_i(x^0)^t d \leq 0, \ i \in I_2 \end{array}\right\} \Rightarrow d^t Hd \geq 0. \qquad \square$$

Definition 1.2

We say that x^0 satisfies the semi-strong second-order sufficiency condition if the following holds :

For any pseudo-optimal subpartition (I_1, I_2) we have

$$\left.\begin{array}{l} d \neq 0, \\ \nabla g_i(x^0)^t d = 0,\ i \in I_1, \\ \nabla g_i(x^0)^t d \leq 0,\ i \in I_2 \end{array}\right\} \Rightarrow d^t H d > 0. \tag{1.8}$$

Let us remark that $((1,p) - I_0^0, I_0^0)$ is a pseudo-optimal subpartition. Hence, (1.8) is stronger than (1.5). On other hand, the semi-strong second-order sufficiency condition is obviously weaker than the strong second-order sufficiency condition. If the problem $(1.1)_0$ is convex (i.e. if f and g_i, $i = m+1, \ldots, p$ are convex and g_i, $i = 1, \ldots, m$ are linear) then $((1,p) - I_0^0, \emptyset)$ is a pseudo-optimal subpartition. Hence, the strong and semi-strong conditions are equivalent for convex problems. We stress the fact that the hypothesis is made on f and g, but is independent of the perturbations.

The paper is organized as follows. In section 2 we only assume that the standard second-order sufficiency condition holds at x_0 and, considering a primal-dual local solution $(x^\varepsilon, \lambda^\varepsilon)$ of $(1.1)_\varepsilon$ we study the limit points of $1/\varepsilon(x^\varepsilon - x^0, \lambda^\varepsilon - \lambda^0)$. We show that the limit points do not necessarily correspond to the local minima of the tangent quadratic problem (an example where this curious phenomenon occurs is given in the Appendix). We call those limit points "pseudo-solutions" of the tangent quadratic problem. In section 3 we study how to compute the local minima associated to a given non-degenerate pseudo-solution of the tangent quadratic problem and we deduce the characterization of the finite set of the local minima of $(1.1)_\varepsilon$, close to x^0, when the semi-strong second-order sufficiency condition holds.

II - SOME RESULTS USING ONLY THE STANDARD SECOND-ORDER SUFFICIENCY CONDITION

As we are interested in the sequel in the directional derivatives of the solutions of $(1.1)_\varepsilon$, we limit our analysis to the case $\varepsilon > 0$. We define a <u>branch</u> to be a mapping $\varepsilon \to (x^\varepsilon, \lambda^\varepsilon)$, defined for ε small enough, such that $(x^\varepsilon, \lambda^\varepsilon) \to (x^0, \lambda^0)$ when $\varepsilon \to 0$. We will speak of a <u>branch of solutions</u> if x^ε is a local solution of $(1.1)_\varepsilon$ and λ^ε is its (uniquely defined for ε small and x^ε close to x^0) associated Lagrange multiplier, i.e.

$$
\begin{aligned}
&\text{(i)} \quad \nabla f^{\varepsilon}(x^{\varepsilon}) + \sum_{i=1}^{p} \lambda_i^{\varepsilon} \nabla g_i^{\varepsilon}(x^{\varepsilon}) = 0, \\
&\text{(ii)} \quad g_i^{\varepsilon}(x^{\varepsilon}) = 0, \ i = 1, \ldots, m, \qquad\qquad (2.1) \\
&\text{(iii)} \quad g_i^{\varepsilon}(x^{\varepsilon}) \le 0, \ \lambda_i^{\varepsilon} \ge 0, \ \lambda_i^{\varepsilon} \, g_i^{\varepsilon}(x^{\varepsilon}) = 0, \ i = m+1, \ldots, p.
\end{aligned}
$$

Set

$$
d^{\varepsilon} = 1/\varepsilon(x^{\varepsilon}-x^{0}), \ \mu^{\varepsilon} = 1/\varepsilon(\lambda^{\varepsilon}-\lambda^{0}).
$$

Under the hypotheses of Theorem 1.1, $(d^{\varepsilon}, \mu^{\varepsilon})$ is bounded. Denote

$$
\begin{aligned}
&c_0 = \partial(\nabla f^{\varepsilon}(x^{0}) + \sum_{i=1}^{p} \lambda_i^{0} \nabla g_i^{\varepsilon}(x^{0}))/\partial\varepsilon(\varepsilon = 0), \\
&c_i = \partial g_i^{\varepsilon}(x^{0})/\partial\varepsilon(\varepsilon = 0), \ i = 1, \ldots, p.
\end{aligned}
$$

The usual linearization techniques applied to the system (2.1) lead to the system (see e.g. [2]) :

$$
\begin{aligned}
&Hd + c_0 + \sum_{i=1}^{p} \mu_i \nabla g_i(x^{0}) = 0, \\
&\nabla g_i(x^{0})^{t} d + c_i = 0, \ i \in (1,p) - I_0^{0}, \qquad\qquad (2.2) \\
&\nabla g_i(x^{0})^{t} d + c_i \le 0, \ \mu_i \ge 0, \ \mu_i(\nabla g_i(x^{0})^{t} d + c_i) = 0, \ i \in I_0^{0}.
\end{aligned}
$$

For any solution (d, μ) of (2.2) we will denote

$$
\begin{aligned}
&J(d) = \{i \in (1,p) \ ; \ \nabla g_i(x^{0})^{t} d + c_i = 0\}, \\
&J_0(d) = \{i \in J(d) \cap I_0^{0} \ ; \ \mu_i = 0\}.
\end{aligned}
$$

Some simple computations involving the difference between (1.3) and (2.1) imply the

Proposition 2.1

We assume that (1.2), (1.3) and (1.5) hold. Let (d,μ) be a limit point of $(d^{\varepsilon},\mu^{\varepsilon})$ when $\varepsilon \to 0$. Then (d,μ) satisfies (2.2). □

Now, let us analyse what supplementary information on (d,μ) is given by the second-order necessary condition satisfied by $(x^\varepsilon,\lambda^\varepsilon)$. We denote

$$I_0^\varepsilon(x^\varepsilon) = \{i = m+1, \ldots, p \; ; \; g_i^\varepsilon(x^\varepsilon) = 0 \text{ and } \lambda_i^\varepsilon = 0\}.$$

The second-order necessary condition for x^ε is

$$\left.\begin{array}{l} \nabla g_i^\varepsilon(x^\varepsilon)^t d = 0, \; i \in I^\varepsilon(x^\varepsilon) - I_0^\varepsilon(x^\varepsilon), \\ \nabla g_i^\varepsilon(x^\varepsilon)^t d \le 0, \; i \in I_0^\varepsilon(x^\varepsilon) \end{array}\right\} \Rightarrow \; d^t H^\varepsilon d \ge 0, \tag{2.3}$$

where

$$H^\varepsilon = \nabla^2 f^\varepsilon(x) + \sum_{i=1}^{p} \lambda_i^\varepsilon \nabla^2 g_i^\varepsilon(x^\varepsilon).$$

<u>Proposition 2.2</u>

Assume that (1.2), (1.3) and (1.5) hold. Let $\{\varepsilon^k\}_{k\in\mathbb{N}}$ be a sequence such that $\varepsilon^k \to 0$ and let $\{x^{\varepsilon^k}\}$ be a sequence of local solutions of $(1.1)_{\varepsilon^k}$ with associated unique Lagrange multiplier λ^{ε^k}. Let (I_1,I_2) be such that $(I_1,I_2) = (I^{\varepsilon^k}(x^{\varepsilon^k}) - I_0^{\varepsilon^k}(x^{\varepsilon^k}), I_0^{\varepsilon^k}(x^{\varepsilon^k}))$.

Then (I_1,I_2) is a pseudo-optimal subpartition, and if (d,μ) is a limit point of $1/\varepsilon^k(x^{\varepsilon^k} - x^0, \lambda^{\varepsilon^k} - \lambda^0)$, then

$$\begin{array}{l} J(d) - J_0(d) \subset I_1 \subset J(d), \\ I_2 \subset J_0(d) - I_1. \end{array} \tag{2.4}$$

<u>Proof</u>

Since (1.2) holds, the condition (2.3) holds at x^{ε^k} and thus

$$\left.\begin{array}{l} \nabla g_i(x^0)^t d = 0,\ i \in I_1, \\ \nabla g_i(x^0)^t d \le 0,\ i \in I_2 \end{array}\right\} \Rightarrow d^t\, Hd \ge 0.$$

Let (d,μ) be a limit point of $(d^{\varepsilon^k}, \mu^{\varepsilon^k}) = 1/\varepsilon^k(x^{\varepsilon^k}-x^0, \lambda^{\varepsilon^k}-\lambda^0)$. We have that $I_1 \cup I_2 \subset J(d)$, and $\lambda_i^{\varepsilon^k} \neq 0$ for ε^k small enough if $i > m$ and $i \in J(d) - J_0(d)$; hence $J(d) - J_0(d) \subset I_1$. As $I_1 \cap I_2 = \emptyset$, we deduce that

$$I_2 \subset J(d)-I_1 = J_0(d)-I_1. \qquad \square$$

It is usual to connect system (2.2) to the tangent quadratic problem

$$\begin{array}{l} \text{Min } c_0^t d + 1/2\ d^t\, Hd\ , \\ \nabla g_i(x^0)^t d + c_i = 0,\ i \in (1,p) - I_0^0, \\ \nabla g_i(x^0)^t d + c_i \le 0,\ i \in I_0^0. \end{array} \qquad \text{(TQP)}$$

It is easy to verify that the standard second-order sufficiency condition (1.5) together with (1.2) imply that (TQP) has at least one local solution.

The first-order optimality conditions of(TQP) are precisely relations (2.2). As (TQP) is not convex, the first-order necessary conditions (2.2) are not sufficient for optimality. Notice however, that as (TQP) is quadratic, the second-order necessary conditions of (TQP) are also sufficient for local optimality.

Proposition 2.2 does not imply that a branch of local solutions of $(1.1)_\varepsilon$ is connected to a solution of the (TQP). In fact we give an example in the Appendix, where a branch of local solution is not connected to a solution of the (TQP), but only to some $d \in \mathbb{R}^n$ such that (2.2) is satisfied for some μ and (2.4) is satisfied for some pseudo-optimal subpartition.

Definition 2.1

We say that d is a pseudo-solution of the tangent quadratic problem associated to (I_1, I_2) if (2.4) is satisfied and there exists some μ such that (d, μ) satisfy (2.2). □

Remark 2.1

(i) A pseudo-solution of the tangent quadratic problem associated to (I_1, I_2) is a local solution of what we may call a pseudo-tangent quadratic problem

$$\begin{aligned} &\min c_0^t d + 1/2\, d^t Hd, \\ &\nabla g_i(x^0)^t d + c_i = 0, \; i \in I_1, \\ &\nabla g_i(x^0)^t d + c_i \leq 0, \; i \in I_2. \end{aligned} \tag{2.5}$$

Problem (2.5) is obtained from the (TQP) by discarding some inequality constraints and by converting some others into equality constraints.

(ii) We remark that a pseudo-solution of the (TQP) may be associated to several pseudo-optimal subpartitions.

(iii) A pseudo-solution of the (TQP) is a stationary point of (TQP), as is satisfies (2.2), but the converse is not always true, as shown by the example of the Appendix. □

III - NON DEGENERATE PSEUDO-SOLUTIONS OF THE TQP AND THE SEMI-STRONG SECOND-ORDER SUFFICIENCY CONDITION

This section contains the main results. We first give some results using a non-degeneracy hypothesis on a given pseudo-solution of the (TQP). Then we characterize the set of solutions of the perturbed problem under the semi-strong second-order sufficiency condition.

Definition 3.1

A pseudo-optimal subpartition (I_1,I_2) is non-degenerate if the following condition holds

$$\left.\begin{array}{l} \delta \neq 0, \\ \nabla g_i(x^0)^t\,\delta = 0,\ i \in I_1, \\ \nabla g_i(x^0)^t\,\delta \leq 0,\ i \in I_2 \end{array}\right\} \Rightarrow \delta^t H\delta > 0. \qquad (3.1)$$

If, in addition, (I_1,I_2) is associated to a pseudo-solution d of the (TQP), we say that d is a non-degenerate pseudo-solution of the (TQP) and that (d,I_1,I_2) is non-degenerate. □

Remark 3.1

There is at most one pseudo-solution of the (TQP) associated to a non-degenerate pseudo-optimal subpartition (I_1,I_2) because it is a part of the solution (d,μ) of the linear system

$$Hd + \sum_{i \in I_1 \cup I_2} \mu_i \nabla g_i(x^0) = -c_0,$$
$$\nabla g_i(x^0)^t d = -c_i,\ i \in I_1 \cup I_2,$$

whose solution is unique, since (1.2) and (3.1) hold. □

We consider the related non linear system

$$\begin{array}{l} \nabla f^\varepsilon(x^\varepsilon) + \sum_{i \in I_1 \cup I_2} \lambda_i^\varepsilon\, \nabla g_i^\varepsilon(x^\varepsilon) = 0, \\ g_i^\varepsilon(x^\varepsilon) = 0,\ i \in I_1 \cup I_2. \end{array} \qquad (3.2)$$

Lemma 3.1

We assume that (1.2), (1.3) and (3.1) hold. Then, for ε small enough, there is a unique C^1 branch $(x^\varepsilon, \lambda^\varepsilon)$ such that (3.2) is satisfied and $\lambda_i^\varepsilon = 0$, $i \in (1,p) - (I_1 \cup I_2)$. □

Proof

System (3.2) is satisfied at $\varepsilon = 0$ by (x^0, λ^0). By remark 3.1, the Jacobian of system (3.2) with respect to $(x, \lambda_i,\ i \in I_1 \cup I_2)$, is non degenerate. Hence, by the implicit function theorem, system (3.2) has, for ε small, a unique solution $(x^\varepsilon,\ \lambda_i^\varepsilon,\ i \in I_1 \cup I_2)$. Taking $\lambda_i^\varepsilon = 0,\quad i \in (1,p) - (I_1 \cup I_2)$, we get the result. □

We will say that this branch is associated to (d, I_1, I_2). The following proposition gives a means to recognize whether this branch corresponds to the solutions of $(1.1)_\varepsilon$.

Proposition 3.1

We assume that (1.2), (1.3) and (3.1) holds. Let $(x^\varepsilon, \lambda^\varepsilon)$ be a branch associated to a non-degenerate triple (d, I_1, I_2). Then, for ε small enough, a necessary condition for x^ε to be a local minimum of $(1.1)_\varepsilon$ is

$$\left.\begin{array}{l} \text{(i)}\quad g_i^\varepsilon(x^\varepsilon) \le 0,\ i \in J_0(d) - (I_1 \cup I_2), \\ \text{(ii)}\ \lambda_i^\varepsilon \ge 0,\ i \in I_0^0 \cap (I_1 \cup I_2), \end{array}\right. \qquad (3.3)$$

$$(I^\varepsilon(x^\varepsilon) - I_0^\varepsilon(x^\varepsilon),\ I_0^\varepsilon(x^\varepsilon)) \text{ is a pseudo-optimal subpartition,} \qquad (3.4)$$

and a sufficient condition is that (3.3) holds and

$$\begin{array}{l}(I^\varepsilon(x^\varepsilon) - I_0^\varepsilon(x^\varepsilon),\ I_0^\varepsilon(x^\varepsilon)) \text{ is a non-degenerate} \\ \text{pseudo-optimal subpartition.}\end{array} \qquad (3.5)$$

□

Proof

The definition of $J(d)$ implies that for ε small, $g_i^\varepsilon(x^\varepsilon) < 0$ if $i \notin J(d)$ and by the definition of that branch and (2.4), $g_i^\varepsilon(x^\varepsilon) = 0$ for i in $J(d) - J_0(d)$ and $I_1 \cup I_2$. Hence x^ε will be feasible for $(1.1)_0$ iff (3.3i) holds. In the same way, we see that as $\lambda_i^\varepsilon = 0$ if $i \in I_1 \cup I_2$, and $\lambda_i^\varepsilon > 0$ if $i \geq m+1$ and $i \in I_0^0$, then λ^ε will be feasible iff (3.3ii) holds. As λ^ε is the only possible multiplier associated to x^ε, condition (3.3) is necessary. If x^ε is a local solution of $(1.1.)_\varepsilon$, the necessity of (3.4) for ε small enough is a consequence of proposition 2.2. Let us now prove that (3.3) and (3.5) are a sufficient condition. If x^ε satisfies (3.3) but is not a local solution, then for some $\delta^\varepsilon \neq 0$:

$$\nabla g_i(x^\varepsilon)^t \delta^\varepsilon = 0, \ i \in I^\varepsilon(x^\varepsilon) - I_0^\varepsilon(x^\varepsilon),$$
$$\nabla g_i(x^\varepsilon)^t \delta^\varepsilon \leq 0, \ i \in I_0^\varepsilon(x^\varepsilon),$$
$$(\delta^\varepsilon)^t H^\varepsilon \delta^\varepsilon \leq 0.$$

We may suppose that $\|\delta^\varepsilon\| = 1$. We take a converging subsequence of such $\{\delta^\varepsilon\}$, when $\varepsilon \to 0$. Passing to the limit in the above relations, we deduce that no limit set of $(I^\varepsilon(x^\varepsilon) - I_0^\varepsilon(x^\varepsilon), I_0^\varepsilon(x^\varepsilon))$ can be a non-degenerate pseudo-optimal subpartition. This proves that (3.3) and (3.5) are a sufficient condition for optimality. □

<u>Corollary 3.1</u>

We assume that the hypotheses of Proposition 3.1 and the semi-strong second-order sufficiency condition (1.8) hold at x^0. Let $(x^\varepsilon, \lambda^\varepsilon)$ be as in Proposition 3.1. Then for ε small enough and x^ε close to x^0, (3.3) and (3.5) are a necessary and sufficient condition for x^ε to be a local solution of $(1.1)_\varepsilon$.
□

<u>Proof</u>

It is sufficient to notice that (1.8) is equivalent to the requirement that any pseudo-optimal subpartition be non degenerate. □

We define

$$X_{\varepsilon,\alpha} = \{x^\varepsilon \in \mathbf{R}^n \ ; \ x^\varepsilon \text{ is a local solution of } (1.1)_\varepsilon \text{ and } \| x^\varepsilon - x^0 \| < \alpha\}.$$

Now, if the semi-strong second-order sufficiency condition holds, we have seen in Section 2 that any branch of solutions of $(1.1)_\varepsilon$ is associated to at least one pseudo-solution of the (TQP) and Corollary 3.1 gives a means to compute the unique solution associated to a given couple of pseudo-solution and associated pseudo-optimal subpartition. As there is a finite number of non-degenerate pseudo-optimal subpartitions, there is also a finite number of such couples. Hence, we have a constructive means to compute all solutions of $(1.1)_\varepsilon$ close to x^0.

Theorem 3.1

We assume hypotheses (1.2), (1.3) and that the semi-strong second-order sufficiency condition (1.8) hold. Let a_1 be as in Theorem 1.1. Then, for α small enough and $\varepsilon < a_1\alpha$, the set $X_{\varepsilon,\alpha}$ is finite and non-empty and, to compute it, it is sufficient to compute the elements of the finite number of branches associated to the pseudo-solutions of the (TQP) and to check whether they satisfy conditions (3.3) and (3.5). □

REFERENCES

1. A.V. FIACCO, **Sensitivity analysis for nonlinear programming using penalty methods.** Math. Prog. 10, pp. 287-311, 1976.

2. K. JITTORNTRUM, **Solution point differentiability without strict complementarity in nonlinear programming.** Math. Prog. Study, 21, pp. 127-138, 1984.

3. S.M. ROBINSON, **Generalized equations and their solutions, Part II : applications to nonlinear programming.** Math. Prog. Study 19, pp. 200-221, 1982.

APPENDIX

Example 1

We give an example in which a branch of solutions of the perturbed problem is associated to a pseudo-solution of the TQP which is not a local solution of the TQP. We consider the following problem

$$\begin{aligned} &\min x_2^2 - x_1^2 + (\varepsilon^2 - 2/3\varepsilon)(2x_1 + x_2) ; \\ &2x_1 - x_2 - \varepsilon \le 0, \\ &-2x_1 - x_2 - \varepsilon \le 0. \end{aligned} \qquad (A1)_\varepsilon$$

For $\varepsilon = 0$ the problem reduces to

$$\min x_2^2 - x_1^2 \;;\; x_2 \ge 2|x_1|. \qquad (A2)$$

It has a unique minimum $x^0 = (0,0)^t$ with an associated unique multiplier $\lambda^0 = (0,0)^t$.

Let us consider the tangent quadratic problem at $\varepsilon = 0$:

$$\begin{aligned} &\min -2/3(2d_1 + d_2) - d_1^2 + d_2^2, \\ &2d_1 - d_2 - 1 \le 0, \\ &- 2d_1 - d_2 - 1 \le 0. \end{aligned} \qquad (A3)$$

The first-order optimality conditions for problem (A3) are

$$\text{(i)}\quad \begin{pmatrix} -2d_1 \\ 2d_2 \end{pmatrix} - \begin{pmatrix} 4/3 \\ 2/3 \end{pmatrix} + \mu_1 \begin{pmatrix} 2 \\ -1 \end{pmatrix} + \mu_2 \begin{pmatrix} -2 \\ -1 \end{pmatrix} = 0,$$

$$\text{(ii)}\quad 2d_1 - d_2 - 1 \leq 0,\ \mu_1 \geq 0,\ \mu_1(2d_1 - d_2 - 1) = 0, \tag{A4}$$

$$\text{(iii)}\quad -2d_1 - d_2 - 1 \leq 0,\ \mu_2 \geq 0,\ \mu_2(-2d_1 - d_2 - 1) = 0.$$

The relations (A4) have two solutions :

- The first for which only the first constraint is binding : we get

$$\bar{d} = (10/9,\ 11/9)^t \ ;\ \bar{\mu} = (16/9,\ 0)^t.$$

- The second for which only the second constraint is binding :

$$\tilde{d} = (-2/3,\ 1/3)^t \ ;\ \tilde{\mu} = (0,0)^t.$$

The first solution $\bar{d}$ is in fact the only local minimum of (A3) and there is a branch of solutions of $(A1)_\varepsilon$ associated to $\bar{d}$:

$$\bar{x}^\varepsilon = (10\varepsilon/9 - 2\varepsilon^2/3,\ 11\varepsilon/9 - 4\varepsilon^2/3)^t \ ;\ \bar{\lambda}^\varepsilon = (16\varepsilon/9 - (5\varepsilon^2/3\ ,\ 0)^t.$$

On the contrary, $\tilde{d}$ is only a saddle point of the criterion of (A3). It is also a pseudo-solution associated to the subpartition $(\{2\},\ \emptyset)$. There is actually a branch of solutions associated to $\tilde{d}$:

$$\tilde{x}^\varepsilon = (-2\varepsilon/3,\ \varepsilon/3)^t \ ;\ \tilde{\lambda}^\varepsilon = (0,\varepsilon^2)^t.$$

Example 2

In this example, the TQP has a stationary point which is not a pseudo-solution. The problem is

$$\min x_2^2 - x_1^2 \ ;\ 2x_1 - x_2 - \varepsilon \leq 0 \ ;\ -2x_1 - x_1 - \varepsilon \leq 0.$$

At $\varepsilon = 0$, the problem still reduces to (A2). The tangent quadratic problem is

$$\min d_1^2 - d_2^2,$$
$$\text{s.t. } 2d_1 - d_2 - 1 \leq 0$$
$$-2d_1 - d_2 - 1 \leq 0.$$

The point $d = (0,0)^t$ is obviously a stationary point of this problem. We prove that it is not a pseudo-solution : $J(d)$ is empty, hence by (2.4) the only subpartition that could be associated to d is $(I_1, I_2) = (\emptyset,\emptyset)$. But the hessian of the lagrangian is not positive definite at d : hence $(\emptyset,\emptyset)$ cannot be a pseudo-optimal subpartition.

AUGMENTABILITY AND EXACT PENALIZABILITY IN NONLINEAR PROGRAMMING UNDER A WEAK SECOND-ORDER SUFFICIENCY CONDITION

Joseph Frédéric BONNANS[#]

ABSTRACT. We prove that some augmented Lagrangian has a strict local minimum at a local solution of a nonlinear programming problem at which some weak second-order sufficiency condition (using F. John multipliers) holds. We also prove that if all multipliers used in the expression of the augmented Lagrangian are Lagrange multipliers, then the exact penalty function is, under a sharp bound on the penalization coefficient, superior to this augmented Lagrangian in the neighbourhood of a local solution.

KEY WORDS. Augmented Lagrangians, non-differentiable penalty functions, nonlinearly constrained optimization.

INRIA, Domaine de Voluceau, BP 105, Rocquencourt, 78153 Le Chesnay Cedex, France

1. INTRODUCTION

We are concerned with the following nonlinear programming problem :

$$\begin{aligned} &\min f(x), \\ &\text{s.t. } g_i(x) = 0,\ i = 1, \ldots, m, \\ &\qquad g_i(x) \le 0,\ i = m+1, \ldots, p, \end{aligned} \tag{1.1}$$

the functions f and g_i, $i = 1, \ldots, p$, being twice differentiable from $\mathbb{R}^n$ to $\mathbb{R}$. A point satisfying all constraints of (1.1) will said to be admissible. Let $\bar{x}$ be a local solution of (1.1). As we are interested by a local analysis we may suppose that $g_i(\bar{x}) = 0$, for $i > m$. It is known (see Ref. 1) that there exists some non-null Fritz John multipliers, i.e. non-null elements $\lambda = (\lambda_0, \ldots, \lambda_p)$ such that

$$\lambda_0 \nabla f(\bar{x}) + \sum_{i=1}^{p} \lambda_i \nabla g_i(\bar{x}) = 0,$$

$$\lambda_0 \ge 0,\ \lambda_i \ge 0,\ i = m+1, \ldots, p.$$

We will say that a Fritz John multiplier λ is a Lagrange multiplier if $\lambda_0 = 1$. A Lagrange multiplier can be identified with an element of $\mathbb{R}^p$. We denote by Λ the set of all Fritz John multipliers. In order to deal with the second-order optimality conditions we define the following closed convex cone :

$$\begin{aligned} C = \{d \in \mathbb{R}^n ;\ &\nabla g_i(\bar{x})^t d = 0,\ i = 1, \ldots, m, \\ &\nabla g_i(\bar{x})^t d \le 0,\ i = m+1, \ldots, p,\ \nabla f(\bar{x})^t d \le 0\}. \end{aligned}$$

We will need the generalized Lagrangian

$$L(x,\lambda) = \lambda_0 (f(x) - f(\bar{x})) + \sum_{i=1}^{p} \lambda_i g_i(x).$$

If λ is a Lagrange multiplier, this Lagrangian reduces to the usual one (up to the constant $-f(\bar{x})$). We will denote by H_λ the hessian with respect to x of L at $(\bar{x},\lambda)$, i.e.

$$H_\lambda = \lambda_0 \nabla^2 f(\bar{x}) + \sum_{i=1}^{p} \lambda_i \nabla^2 g_i(\bar{x}).$$

Under the following condition :

There exists a Lagrange multiplier λ such that

$$d \in C,\ d \neq 0 \Rightarrow d^t H_\lambda d > 0 \tag{1.2}$$

it has been proved (see Ref. 5,6) that some augmented Lagrangian has a strict minimum at $\bar{x}$.

Define $\tilde{g}(x)$ by

$$\tilde{g}_i(x) = \begin{cases} g_i(x) \text{ if } i \leq m, \\ g_i(x)^+ \text{ if not.} \end{cases}$$

Let us define the so-called exact penalty function

$$\theta_r(x) = f(x) + r\|\tilde{g}(x)\|.$$

It has been proved (Ref. 4) that if condition (1.2) holds, and if r is superior to the dual norm of λ, then θ_r has a strict minimum at $\bar{x}$.

Our concern is to generalize these results to the case when (1.2) does not holds. In fact, our main results will need the following hypotheses : one Lagrange multiplier exists and

For any non-null d in C, there exists a Fritz John multipliers λ

such that $d^t H_\lambda d > 0$. (1.3)

This condition is known to be sufficient for $\bar{x}$ to be a strict local minimum (Ref. 1,4). If we change the strict inequality of (1.3) into an inequality we obtain a necessary condition (Ref. 1,7). Hence condition (1.3) is rather weak.

The augmented Lagrangians used in this paper involve the maximization of the Lagrangian over a set of multipliers. Such kind of augmented Lagrangian have already been considered by A.D. Ioffe (Ref. 7). Our results, however, are not covered by those of (Ref. 7), as discussed in the following section. The main result on augmented Lagrangians is that if (1.3) holds and if at least one Lagrange multiplier exists, it is possible to build a continuous augmented Lagrangian, having a strict minimum at the local solution, which is an extension of those considered in (Ref. 6,8).

We also prove that in the neighbourhood of a local solution (and in fact of any admissible point) the non-differentiable penalty function of the type considered above dominates some augmented Lagrangian. From this we deduce in particular that if the multipliers involved in (1.3) are Lagrange multipliers, this penalty function has a strict minimum at the local solution ; this generalizes some results of (Ref. 4).

2. AUGMENTABILITY OF NONLINEAR PROGRAMS

In order to define an augmented Lagrangian, we consider the subsets $\Lambda^{\#}$ of $\mathbb{R}^{p+1}$ whose elements λ are denoted $(\lambda_0, \ldots, \lambda_p)$ having the following property :

$$\Lambda^{\#} \text{ is a compact subset of } \Lambda \text{ and}$$
$$d \in C,\ d \neq 0 \Rightarrow d^t H_\lambda d > 0 \text{ for some } \lambda \text{ in } \Lambda^{\#}. \tag{2.1}$$

The following simple Lemma is close to Theorem 7 of A.D. Ioffe (Ref. 7).

Lemma 2.1 Let Λ^* be a closed subset of Λ such that

$$d \in C,\ d \neq 0 \Rightarrow d^t H_\lambda d > 0 \text{ for some } \lambda \text{ in } \Lambda^*. \tag{2.2}$$

Then there exists $\Lambda^{\#} \subset \Lambda^*$ satisfying (2.1). □

Proof If the conclusion if false, then in particular it does not hold for the following set

$$\Lambda_r^* = \{\lambda \in \Lambda^* \ ; \|\lambda\| \leq r\},$$

for any positive r. Equivalently, there exists a sequence $\{d^k\}$ in C, with $d^k \neq 0$ for any k, such that

$$\lambda \in \Lambda^*, \ \|\lambda\| \leq k \Rightarrow (d^k)^t H_\lambda d^k \leq 0.$$

Let d be a limit point of $d^k/\|d^k\|$. By (2.2), there exists $\lambda°$ in λ^* such that $d^t H_{\lambda°} d > 0$. Consequently, for k great enough, we have

$$(d^k)^t H_{\lambda°} d^k > \|d^k\|^2 \ d^t H_{\lambda°} d/2 > 0,$$

which gives a contradiction. □

We also need two subset I^+, $I°$ of (1,p) having the following property :

$$\left.\begin{aligned} & I^+ \cap I° = \emptyset, \\ & C = \{d \in \mathbb{R}^n \ ; \ \nabla g_i(\bar{x})^t d = 0, \ i \in I^+ \ ; \ \nabla g_i(\bar{x})^t d \leq 0, \ i \in I°\}. \end{aligned}\right\} \quad (2.3)$$

These sets are connected to the Lagrange multipliers, as shown by the (well known) following Lemma.

Lemma 2.2 Let λ be a Lagrange multiplier. Then the sets

$$I°(\lambda) = \{i > m \ ; \ \lambda_i = 0\},$$
$$I^+(\lambda) = (1,p) - I°(\lambda),$$

verify (2.3). □

Proof If λ is a Lagrange multiplier, the following equality holds :

$$\nabla f(\bar{x}) + \sum_{i=1}^{p} \lambda_i \nabla g_i(\bar{x}) = 0.$$

Hence for all d in C, we have

$$0 \geq \nabla f(\bar{x})^t d = - \sum_{i=1}^{p} \lambda_i \nabla g_i(\bar{x})^t d \geq 0.$$

Hence $\nabla f(\bar{x})^t d = 0$, and $\lambda_i \nabla g_i(\bar{x})^t d = 0$, $i = 1, \ldots, p$. This implies that $\nabla g_i(\bar{x})^t d = 0$ for all i in $I^+(\lambda)$ and $\nabla g_i(\bar{x})^t d \leq 0$ for all i in $I°(\lambda)$. The converse is immediate. □

Let $\Lambda^{\#}$ be a closed bounded subset of $\mathbb{R}^p$ and c be a non-negative number. We consider the following augmented Lagrangian

$$L_c(x,\Lambda^{\#},I^+,I^\circ) = \max\{L(x,\lambda),\lambda \in \Lambda^{\#}\} + c(\sum_{i\in I^+} g_i(x)^2 + \sum_{i\in I^\circ} (g_i(x)^+)^2).$$

This Lagrangian verifies, for any $\Lambda^{\#}$, I^+, I° :

$$L_c(\bar{x},\Lambda^{\#},I^+,I^\circ) = 0.$$

Remark 2.1 If $\Lambda^{\#}$ reduces to $\{\lambda^{\#}\}$, $\lambda^{\#}$ being a Lagrange multiplier, and if $(I^+, I^\circ) = (I^+(\lambda^{\#}), I^\circ(\lambda^{\#}))$, this augmented Lagrangian reduces to those considered in M.R. Hestenes (Ref. 6) and is equal, near an admissible point, to those considered in R.T. Rockafellar (Ref. 8).

Definition 2.1 Problem (1.1) is said to be (strictly) augmentable near $\bar{x}$ if there exists a closed bounded subset $\Lambda^{\#} \subset \Lambda$ and (I^+,I°) verifying (2.3) such that $L_c(x,\Lambda^{\#},I^+,I^\circ)$ has a (strict) minimum at $\bar{x}$. □

We state our main result on augmented Lagrangians.

Theorem 2.1 Let $\bar{x}$ be a local solution of (1.1) at which hypothesis (1.3) holds. Let $\Lambda^{\#}$ be such that (2.1) holds and (I^+,I°) be such that (2.3) holds. Then problem (1.1) is strictly augmentable near x and there exist $\alpha > 0$ and $c_0 \geq 0$ such that, for any $c \geq c_0$, the following inequality holds in a neighbourhood of $\bar{x}$:

$$\alpha\|x-\bar{x}\|^2 \leq L_c(x,\Lambda^{\#},I^+,I^\circ). \qquad \square$$

Remark 2.2 The idea of defining the augmented Lagrangian by taking a maximum over a bounded subset of multipliers can be found in A.D. Ioffe (Ref. 7). The main difference between the results of (Ref. 7) and theorem 2.1 is that we do not need hypothesis b) of (Ref. 7, Thm. 10, p. 286). Also, the method of proof is completely different and we give a lower estimate of $L_c(x,\Lambda^{\#},I^+,I^\circ)$. □

In order to prove Theorem 2.1, we need a preliminary result on positively homogeneous forms.

Definition 2.2 A positively homogeneous form P of degree r, $r > 0$, from $\mathbb{R}^n$ onto $\mathbb{R}$ is a mapping such that

$$P(td) = t^r P(d)$$

for all non-negative t. □

Lemma 2.3 Let P,Q be two lower-semi-continuous positively homogeneous forms of degree r, $r > 0$, such that

(i) $Q \geq 0$,
(ii) $P(d) > 0$, for all non null d such that $Q(d) = 0$.

Then there exists $c_1 > 0$ and c_2 such that, for all $c > c_2$ and d in $\mathbb{R}^n$:

$$P(d) + c\,Q(d) \geq c_1 \|d\|^r. \quad \square$$

Proof Suppose that the conclusion does not hold. Then for all $\varepsilon > 0$, there exists a sequence $\{d_\varepsilon^n\}$ such that

$$P(d_\varepsilon^n) + nQ(d_\varepsilon^n) < \varepsilon \|d_\varepsilon^n\|^r.$$

From the definition of positively homogeneous forms we deduce that $P(0)=Q(0)=0$. Hence the above inequality implies that $d_\varepsilon^n \neq 0$. Using the homogeneity of P and Q, we see that

$$P(d_\varepsilon^n/\|d_\varepsilon^n\|) + nQ(d_\varepsilon^n/\|d_\varepsilon^n\|) < \varepsilon.$$

Let d_ε be a limit point of $d_\varepsilon^n/\|d_\varepsilon^n\|$. Then $\|d_\varepsilon\| = 1$ and from the positivity of Q and the lower-semi-continuity of P and Q we deduce that $Q(d_\varepsilon) = 0$ and $P(d_\varepsilon) \leq \varepsilon$. Let d be a limit-point of $\{d_\varepsilon\}$ when $\varepsilon \to 0$. Then $\|d\| = 1$ and, by the positivity of Q and the lower-semi-continuity of P and Q, $Q(d) = 0$ and $P(d) \leq 0$. This gives a contradiction with the hypotheses. □

We remark that Lemma 2.3 is an extension of a result of M.R. Hestenes (Ref. 6) on quadratic forms. The spirit of the proof is itself similar to the one of (Ref. 6).

We now define ($\Lambda^{\#}$, I^{+}, I^{0} being as in theorem 2.1) :

$$\bar{P}(d) = \tfrac{1}{2}\max\{d^t H_\lambda d,\ \lambda \in \Lambda^{\#}\},$$

$$\bar{Q}(d) = \sum_{i\in I^+} (\nabla g_i(\bar{x})^t d)^2 + \sum_{i\in I^0} ((\nabla g_i(\bar{x})^t d)^+)^2.$$

Proposition 2.1 Under the hypotheses of theorem 2.1, there exists $c_3 > 0$ and c_4 such that, for any $c \geq c_4$ and d in $\mathbb{R}^n$:

$$\bar{P}(d) + c\,\bar{Q}(d) \geq c_3\|d\|^2. \qquad \square$$

Proof We use lemma 2.3 with $P = \bar{P}$, $Q = \bar{Q}$ and $r = 2$. As $\Lambda^{\#}$ is compact and $d^t H_\lambda d$ is continuous with respect to d and λ, $\bar{P}$ is continuous. Obviously, $\bar{Q}$ is non negative and continuous, and $\bar{P}$ and $\bar{Q}$ are positively homogeneous forms of degree 2. By (2.3), $\bar{Q}(d) = 0$ iff d is in C ; hence by (2.1), point (ii) of lemma 2.3 holds. The result follows. $\square$

We are now able to state the

Proof of theorem 2.1 As f and g are twice differentiable at $\bar{x}$, we have

$$\lambda_0(f(\bar{x}+d) - f(\bar{x})) = \lambda_0\nabla f(\bar{x})^t d + \lambda_0 d^t\nabla^2 f(\bar{x})d/2 + \lambda_0\, o(\|d\|^2),$$

$$\lambda_i g_i(\bar{x}+d) = \lambda_i\nabla g_i(\bar{x})^t d + \lambda_i d^t\nabla^2 g_i(\bar{x})d/2 + o(\|d\|^2),$$

hence, for all Fritz-John multipliers :

$$L(\bar{x}+d, \lambda) = d^t H_\lambda d/2 + \sum_{i=0}^{p} \lambda_i\, o(\|d\|^2),$$

the terms $o(\|d\|^2)$ being independent on λ. As $\Lambda^{\#}$ is compact, we deduce that

$$\max\{L(\bar{x}+d, \lambda),\ \lambda \in \Lambda^{\#}\} = \bar{P}(d) + o(\|d\|^2).$$

We also have

$$\sum_{i\in I^+} g_i(\bar{x}+d)^2 + \sum_{i\in I^\circ} (g_i(\bar{x}+d)^+)^2 = \bar{Q}(d) + o(\|d\|^2).$$

Finally, we get

$$L_c(\bar{x}+d,\ \Lambda^{\#},\ I^+,\ I^\circ) = \bar{P}(d) + c\bar{Q}(d) + o(\|d\|^2).$$

Using proposition 2.1, we get the desired result. □

Remark 2.3 The usual way in proving that the problem is augmentable under the second-order sufficiency condition is indirect : it consists in using slack variables in order to formulate an equivalent problem with no inequality constraints other than bounds, then to explicitely minimize the augmented Lagrangian with respect to these slack variables (see Ref. 2,6). On the contrary the method given here, based on Lemma 2.3, is direct. □

3. EXACT PENALIZABILITY

Let $\tilde{g}(x)$ be defined as in the introduction. This section is concerned with the exact (non differentiable) penalty function

$$\theta_r(x) = f(x) + r\|\tilde{g}(x)\|.$$

Here $r > 0$ is the penalty coefficient and $\|.\|$ is any norm of $\mathbb{R}^p$. (We will also denote by $\|.\|$ a norm of $\mathbb{R}^n$ when there is no ambiguity). Our aim is to prove that θ_r has, under some convenient hypotheses, a strict minimum at $\bar{x}$ and also that θ_r is superior to the following augmented Lagrangian, in which $\Lambda^{\#}$ is a compact subset of $\mathbb{R}^p$ ($I^+(\lambda)$ and $I^\circ(\lambda)$ are defined as in Lemma 2.2) :

$$L_c^*(x,\Lambda^{\#}) = f(x) + \max_{\lambda\in\Lambda^{\#}} [\lambda^t g(x) + c\sum_{i\in I^+(\lambda)} g_i(x)^2 + c\sum_{i\in I^\circ(\lambda)} (g_i(x)^+)^2].$$

Obviously $L_c^*(\bar{x},\Lambda^{\#})$ is equal to $f(\bar{x})$. We first give an augmentability property concerning $L_c^*(x,\ \Lambda^{\#})$.

Theorem 3.1 Let $\Lambda^{\#}$ be a set of Lagrange multipliers satisfying (2.1). Then there exists $c_s \geq 0$ such that, for any $c \geq c_s$, the following inequality holds in a neighbourhood of $\bar{x}$:

$$L_c^*(\bar{x},\Lambda^{\#}) + c\|x-\bar{x}\|^2 \leq L_c^*(x,\Lambda^{\#}). \quad \square \tag{3.1}$$

Proof As all elements of $\Lambda^{\#}$ are Lagrange multipliers, we have

$$L_c^*(x,\Lambda^{\#}) - f(\bar{x}) = \max_{\lambda\in\Lambda^{\#}} [L(x,\lambda) + c\sum_{i\in I^+(\lambda)} g_i(x)^2 + c\sum_{i\in I^\circ(\lambda)} (g_i(x)^+)^2]$$

$$\geq \max_{\lambda\in\Lambda^{\#}} L(x,\lambda) + c \min_{\mu\in\Lambda^{\#}} [\sum_{i\in I^+(\mu)} g_i(x)^2 + \sum_{i\in I^\circ(\mu)} (g_i(x)^+)^2]$$

$$= \min_{\mu\in\Lambda^{\#}} L_c(x,\Lambda^{\#},I^+(\mu), I^\circ(\mu)).$$

By Lemma 2.2 and Theorem 2.1, for each μ in $\Lambda^{\#}$, there exists $\alpha > 0$ depending on $I^+(\lambda)$ and $I^\circ(\lambda)$ such that

$$\alpha\|x-\bar{x}\|^2 \leq L_c(x,\Lambda^{\#}, I^+(\mu), I^\circ(\mu)).$$

As there is a finite number of couples $(I^+(\mu), I^\circ(\mu))$, for some $\beta > 0$, the following inequality holds near $\bar{x}$:

$$\beta\|x-\bar{x}\|^2 \leq \min_{\mu\in\Lambda^{\#}} L_c(x,\Lambda^{\#},I^+(\mu), I^\circ(\mu)).$$

Hence (3.1) holds. $\square$

We now prove that for r great enough, θ_r dominates L_c^* in the neighbourhood of any admissible point. For this we need the concept of dual norm : the dual norm $\|\cdot\|_D$ of $\|\ \|$ is defined by

$$\|z\|_D = \max \{z^t y \ ; \ \|y\| \leq 1\}.$$

An immediate consequence of the definition is the generalized Cauchy inequality

$$|z^t y| \leq \|z\|_D \|y\|, \text{ for all } z,y \text{ in } \mathbb{R}^p.$$

Theorem 3.2 Let x° be an admissible point and $\Lambda^{\#}$ be any bounded subset of $\mathbb{R}^p$. Then, if

$$r > \sup \{\|\lambda\|_D,\ \lambda \in \Lambda^{\#}\}, \tag{3.2}$$

the following inequality holds in a neighbourhood of x° :

$$L_c^*(x^\circ, \Lambda^{\#}) \leq \theta_r(x). \qquad \square$$

Proof We have

$$\theta_r(x) - L_c^*(x, \Lambda^{\#}) = r\|\tilde{g}(x)\| - \max_{\lambda \in \Lambda^{\#}} [\lambda^t g(x) + c[\sum_{i \in I^+(\lambda)} g_i(x)^2 + \\ + \sum_{i \in I^\circ(\lambda)} (g_i(x)^+)^2]].$$

Hence it is necessary and sufficiency to prove for each λ in $\Lambda^{\#}$ the positivity near x° of

$$\Delta = r\|\tilde{g}(x)\| - [\lambda^t g(x) + c \sum_{i \in I^+(\lambda)} g_i(x)^2 + c \sum_{i \in I^\circ(\lambda)} g_i(x)^+)^2].$$

Let us define $g^{\#}(x)$ by

$$g_i^{\#} = \begin{cases} g_i(x) \text{ if } i \leq m \text{ or } \lambda_i > 0, \\ g_i(x)^+ \text{ if not.} \end{cases}$$

Then, denoting by $\|.\|_2$ the Euclidean norm in $\mathbb{R}^p$, we have for some $\alpha > 0$:

$$\begin{aligned} \Delta &= r\|\tilde{g}(x)\| - \lambda^t g(x) - c\|g^{\#}(x)\|_2^2, \\ &\geq r\|\tilde{g}(x)\| - \lambda^t g(x) - 2c[\|\tilde{g}(x)\|_2^2 + \|\tilde{g}(x) - g^{\#}(x)\|_2^2] \\ &\geq r\|\tilde{g}(x)\| - \lambda^t g(x) - \alpha(\|\tilde{g}(x)\|^2 + \|\tilde{g}(x) - g^{\#}(x)\|^2). \end{aligned}$$

We may write that

$$- \lambda^t g(x) = - \lambda^t \tilde{g}(x) + \lambda^t(\tilde{g}(x) - g(x))$$

$$\geq - \|\lambda\|_D \|\tilde{g}(x)\| + \lambda^t(\tilde{g}(x) - g(x)).$$

Hence

$$\Delta \geq (r - \|\lambda\|_D - \alpha\|\tilde{g}(x)\|) \|\tilde{g}(x)\| + \lambda^t(\tilde{g}(x) - g(x)) - \alpha\|\tilde{g}(x) - g^{\#}(x)\|^2.$$

As $\lambda^t g(x) = \lambda^t g^{\#}(x)$, using 3.2, we deduce that, in a neighbourhood of x° :

$$\Delta \geq \lambda^t(\tilde{g}(x) - g^{\#}(x)) - \alpha\|\tilde{g}(x) - g^{\#}(x)\|^2.$$

Recalling the definition of $\tilde{g}$ and $g^{\#}$, we see that for some $\beta > 0$:

$$\lambda^t(\tilde{g}(x) - g^{\#}(x)) = \sum_{\substack{i>m \\ \lambda_i \neq 0}} \lambda_i(g_i^+(x) - g_i(x)) = \geq \beta\|\tilde{g}(x) - g^{\#}(x)\|.$$

Hence

$$\Delta \geq (\beta - \alpha\|\tilde{g}(x) - g^{\#}(x)\|) \|\tilde{g}(x) - g^{\#}(x)\|,$$

which gives the desired result. □

We now combine Theorems 2.1, 3.1 and 3.2 in order prove that θ_r may have a strict minimum at a local solution.

Theorem 3.3 Under the hypothesis of theorem 2.1, if all element of $\Lambda^{\#}$ are Lagrange multipliers, and if inequality (3.2) holds, θ_r has a strict local minimum at $\bar{x}$ and the following inequality

$$f(\bar{x}) + \alpha\|x - \bar{x}\|^2 \leq \theta_r(x)$$

holds for some $\alpha > 0$ in a neighbourhood of $\bar{x}$. □

Remark 3.1

(i) By Theorem 3.2, we can say that augmentability (in the sense that L_c^* has a local minimum at $\bar{x}$) implies penalizability.

(ii) If $\Lambda^{\#} = \{\lambda\}$ for some Lagrange multiplier λ in $\mathbb{R}^p$, our result reduces to one of S.P. Han, O.L. Mangasarian (Ref. 4, Thm 4.6, p. 265), with the same bound on the penalty coefficient. □

REFERENCES

1. A. BEN-TAL : Second order and related extremality conditions in nonlinear programming. J. Opt. Th. Appl. 31, 143-165, (1980).

2. D.P. BERTSEKAS : Constrained optimization and Lagrange multiplier methods. Academic Press, New York, 1982.

3. S.P. HAN : A global convergent method for nonlinear programming. J. Opt. Th. Appl. 22, 297-309, (1977).

4. S.P. HAN - O.L. MANGASARIAN : Exact penalty functions in nonlinear programming. Math. Prog. 17(3), 251-269, (1979).

5. M.R. HESTENES : Multiplier and gradient methods. J. Optim. Th. Appl. 4, 303-320, (1969).

6. M.R. HESTENES : Optimization theory, the finite dimensional case. J. Wiley, New York, 1975.

7. A.D. IOFFE : Necessary and sufficient conditions for a local minimum 3 : second order conditions and augmented duality. SIAM J. Cont. Opt. 17, 266-288, (1979).

8. R.T. ROCKAFELLAR : Augmented Lagrange multiplier functions and duality in non convex programming. SIAM J. Control. 12, 268-285, 1974.

Imprimé en France

par

l'Institut National de Recherche en Informatique et en Automatique